Dimerization of Amino Acid Side Chains

de Jong, Djurre H.; Periole, Xavier; Marrink, Siewert J.

Published in:
Journal of Chemical Theory and Computation

DOI:
10.1021/ct200599d

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Dimerization of Amino Acid Side Chains: Lessons from the Comparison of Different Force Fields

Djurre H. de Jong, Xavier Periole, and Siewert J. Marrink*

Groningen Biomolecular Sciences and Biotechnology Institute & Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

Supporting Information

ABSTRACT: The interactions between amino acid side chains govern protein secondary, tertiary, and quaternary structure formation. For molecular modeling approaches to be able to realistically describe these phenomena, the underlying force fields have to represent these interactions as accurately as possible. Here, we compare the side chain−side chain interactions for a number of commonly used force fields, namely the all-atom OPLS, the united-atom GROMOS, and the coarse-grain MARTINI force field. We do so by calculating the dimerization free energies between selected pairs of side chains and structural characterization of their binding modes. To mimic both polar and nonpolar environments, the simulations are performed in water, *n*-octanol, and decane. In general, reasonable correlations are found between all three force fields, with deviations on the order of 1 kT in aqueous solvent. In apolar solvent, however, significantly larger differences are found, especially for charged amino acid pairs between the OPLS and GROMOS force fields, and for polar interactions in the MARTINI force field in comparison to the higher resolution models. Interestingly, even in cases where the dimerization free energies are similar, the binding mode may differ substantially between the force fields. This was found to be especially the case for aromatic residues. In addition to the inter-force-field comparison, we compared the various force fields to a knowledge-based potential. The two independent approaches show good correlation in aqueous solvent with an exception of aromatic residues for which the interaction strength is lower in the knowledge-based potentials.

1. INTRODUCTION

For the past couple of decades, molecular dynamics (MD) computer simulations have become a valuable tool to study proteins, from their dynamics and folding into a functional structure to their assembly into complexes.[1−10] The correct formation of these structures critically depends on a delicate balance between attractive and repulsive forces of the constituent amino acids. Together, these interactions determine the secondary, tertiary, and quaternary structures formed, giving rise to biological function. For molecular modeling approaches to be successful, it is therefore essential to describe these interactions as accurately as possible. It is a continuous challenge to improve the molecular force fields, in order to capture more and more realistic behavior. Much effort has recently been devoted to the improvement of the protein backbone conformational preferences, which has an important contribution to protein folding.[11−13] Tests on the performance of the side chain interactions, however, are less abundant. In this work, we focus on the performance of side chains (SC) by looking at association constants and kinetics of SC amino acid analogues (SCA). Three different force fields (FF) are compared: two atomistic, the OPLS and GROMOS 53, and one coarse-grain (CG), MARTINI 2.1.

The parametrization of the SC nonbonded interactions in MD force fields might follow different strategies, leading to different accuracies and resolutions. The OPLS FF has been parametrized against conformational energies and properties of organic liquids.[14,15] In this FF, hydrogen atoms do not carry van der Waals interactions but are represented by a Coulomb interaction site. The GROMOS 53A5/53A6 FF has been parametrized against free enthalpies of solvation of small molecular building blocks in water and cyclohexane.[16] GROMOS is a united-atom (UA) FF where aliphatic hydrogens are not explicitly modeled; i.e., they have no nonbonded interactions. This reduces the number of interaction sites to compute and thus leads to a speed up of simulations. The CG MARTINI FF has even fewer interaction sites: on average, four heavy atoms are mapped into one bead. This leads to a significant computational speedup, but at the cost of simulation details. The MARTINI FF has been parametrized primarily targeting the partitioning free energies of model compounds between polar and apolar phases.[17,18]

To assess amino acid SC parameters, we report on the interaction strength of SCA as characterized by their dimerization (or binding, association) free energies, $\Delta G^{\mathrm{dimer}}$. Earlier studies looked at the dimerization free energies of a limited set of SC pairs in water,[19] or at all SC combinations in water in order to derive effective residue-based contact energies for the development of CG potentials.[20] Here, we systematically calculated $\Delta G^{\mathrm{dimer}}$ for pairs of SCA in polar (water) as well as apolar solvents (*n*-octanol and decane) in order to probe the SC in a mimic of the diversity of biological environments. While for soluble proteins the values of $\Delta G^{\mathrm{dimer}}$ obtained in water will be most relevant, in lipid membranes the dielectric screening much more resembles that of apolar solvents, and consequently the values obtained in octanol and decane are more relevant. For all solvents, we analyzed the structural binding modes found in the different force fields and looked at the kinetics of the

Received: August 25, 2011
Published: February 8, 2012

association process through calculation of the survival correlation times of the SCA pairs, in order to better explain possible differences in ΔG^{dimer}.

The comparison of different force fields can only address the consistency between the force fields. While a strong similarity between force fields that have been parametrized against different experimental data would certainly increase the confidence in the force fields, only a comparison to experimental data can give a direct assessment of the quality of the force fields. To the best of our knowledge, experimental binding free energies of SCAs in solution are not available in the literature. The only data we found pertain to the backbone analogue.[21] However, an indirect measure of the affinity of SCAs for each other can be obtained from a population analysis of their contacts in proteins, which has been used to estimate effective binding free energies.[22] Thus, as an independent check for the side chain interactions, we compare the results obtained for the molecular dynamics FFs to a knowledge-based potential.

The article is organized as follows. In section 2, a description of the FF parameters and methods to obtain ΔG^{dimer} is given. In section 3, ΔG^{dimer} values for 21 different amino acid SCA pairs are reported and compared. This section is split into the comparison of the high-resolution FFs, OPLS, and GROMOS (section 3.1), followed by the comparison of the MARTINI CG FF to both OPLS and GROMOS FF (section 3.2), the analysis of structural binding modes (section 3.3), and finally the comparison of all three FFs to a knowledge-based potential (section 3.4). In section 3.5, we analyze the dimerization kinetics of the different FFs in water. A short concluding section ends this paper.

2. METHODS

2.1. System Setup. All systems simulated consisted of two amino acid SCAs solvated in either water, *n*-octanol, or decane. The systems contained 880, 113, and 80 solvent molecules for water, octanol, and decane, respectively, in a cubic unit cell of ~3.0 nm. For the CG system, 880 water molecules corresponds to 220 water beads, of which 10% were replaced by antifreeze particles.[17]

The dimerization free energies, ΔG^{dimer}, were calculated for 21 selected amino acid SCA pairs (see Tables 1 and 2) and were classified into five groups: charged (Lys−Glu, Lys−Lys, Arg−Asp), polar (Asn−Asn, Ser−Ser, Gln−Asn, Gln−Gln), apolar (Leu−Leu, Val−Val, Leu−Val), mixed (Leu−Asn, Leu−Gln, Leu−Lys, Asn−Lys, Met−Ser), and aromatic (Trp−Trp, Tyr−Tyr, Phe−Phe, His−Phe, Trp−Tyr). The SCAs were constructed by replacing the C_α in the C_β−C_α bond by a hydrogen and omitting the rest of the backbone. Backbone−backbone dimerization free energies were also calculated. A backbone (BB) analogue was defined by a Gly residue as found in a continuous protein backbone. In other words, the C-terminus only has a carbonyl oxygen attached, not a complete acid group. Although this is not a chemically realistic and stable molecule, it best mimics the interactions found in a protein.

Table 1. Three-Letter Amino Acid Abbreviations and Side Chain Analogs Used

side chain	analogue
Leu	*iso*-butane
Val	propane
Lys	butyl amine
Glu	propionic acid
Phe	toluene
Asn	acetamide
Trp	3-methylindole
Tyr	*p*-cresol
Ser	methanol
Met	methyl ethyl sulfide
Arg	propyl guanidine
Asp	acetic acid
Gln	propion amide
His	4-methyl imidazole
BB	*see text*

The SCAs were initially randomly placed in the solvent box. The systems were then minimized and equilibrated for 20 000 steps, followed by a 240 ns production run during which frames were saved every 500 steps for analysis. All simulations were performed using version 4.0.x of the GROMACS[23] simulation package.

2.2. Simulation and Force Field Parameters. All simulations were performed using three levels of description: the all-atom (AA), the united-atom (UA), and the coarse-grain (CG) level. For the AA simulations, the OPLS FF[14,15] was used, as implemented in the GROMACS 4.0.x package.[23] The SPC model[24] was used to represent water. Although the TIP3P model is a more common choice in combination with OPLS, we decided upon SPC, to ensure that we were comparing properties of the protein FFs and not variations of the solvents. Parameters for *n*-octanol and decane compatible with the OPLS FF were taken from Garrido et al.[25] The Lennard-Jones (LJ) and Coulombic interactions were cutoff at 1.4 nm with a neighbor list update every five steps. To correct for the truncation of electrostatic interactions beyond the cutoff, a reaction-field (RF) correction was applied.[26] The relative dielectric constant for the RF, ε_{rf}, was set to 54, 8.8, and 2 for water, *n*-octanol, and decane, respectively.

For the UA simulations, the GROMOS FF[16] was used, as implemented in the GROMACS 4.0.x package.[23] For version 53 of the GROMOS FFs, two parameter sets were used. For the apolar solvents, 53A5 was used, for which the amino acid parameters have been derived in cyclohexane. For polar solvents, 53A6 was used, which is the same as 53A5, except for the partial charges, which have been adjusted to reproduce the hydration free enthalpies in water. The SPC model[24] was used to represent water. Parameters for *n*-octanol and decane compatible with the GROMOS FF were taken from Garrido et al.[25] The LJ interactions and Coulombic interactions were handled as done in the AA simulations. The relative dielectric constant for the RF, ε_{rf}, was set to 54, 8.8, and 2 for water, *n*-octanol, and decane, respectively.

For the CG simulations, the MARTINI FF[17] and its extension to proteins[18] were used. The parameters for the SCAs and the solvents were as reported in these publications. Simulations were also performed with the recently parametrized polarizable MARTINI water model.[27] For the backbone analogue, the "P5" particle type was used, which represents the backbone in an unstructured peptide. Following the standard protocol for MARTINI, the LJ interactions were smoothly shifted to zero between 0.9 and 1.2 nm, using the switch potential implemented in GROMACS. The Coulombic interactions were shifted between 0 and 1.2 nm, for both regular and polarizable MARTINI. The neighbor list was updated every 10 steps. Effective dielectric constants $\varepsilon_r = 15$ for regular MARTINI and $\varepsilon_r = 2.5$ for polarizable MARTINI water were used to screen the Coulombic interactions.

Table 2. Binding Free Energies (k_BT) for Pairs of Amino Acids in Water, Octanol, and Decane[a]

	water				octanol			decane		
pair	pol. MART.	MARTINI	GROMOS	OPLS	MARTINI	GROMOS	OPLS	MARTINI	GROMOS	OPLS
charged										
Lys−Glu	0.00 ± 0.13	−1.67 ± 0.05	−0.08 ± 0.08	−1.35 ± 0.06	−3.53 ± 0.06			−10.35 ± 1.71		
Lys−Lys	0.37 ± 0.12	0.85 ± 0.18	0.88 ± 0.22	1.21 ± 0.31	0.20 ± 0.12			−5.14 ± 0.07		
Arg−Asp	−0.16 ± 0.12	−1.66 ± 0.05	−0.13 ± 0.07	−2.78 ± 0.05	−4.16 ± 0.06					
polar										
Ser−Ser	−0.02 ± 0.10	0.21 ± 0.12	0.58 ± 0.15	0.54 ± 0.14	−0.24 ± 0.09	−0.17 ± 0.15	−0.33 ± 0.15	−0.90 ± 0.06	−4.61 ± 0.14	−2.33 ± 0.06
Gln−Asn	0.11 ± 0.12	0.13 ± 0.14	0.07 ± 0.11	−0.15 ± 0.10	−0.82 ± 0.06	0.00 ± 0.13	0.53 ± 0.21	−1.65 ± 0.12	−7.32 ± 0.74	
Asn−Asn	0.23 ± 0.13	0.35 ± 0.15	0.22 ± 0.11	0.06 ± 0.11	−0.81 ± 0.07	0.11 ± 0.14	−0.53 ± 0.09	−1.50 ± 0.10	−7.14 ± 1.13	
Gln−Gln	0.25 ± 0.13	−0.02 ± 0.13	−0.14 ± 0.10	−0.53 ± 0.11	−0.71 ± 0.07	−0.47 ± 0.11	−0.55 ± 0.12	−1.34 ± 0.05	−6.96 ± 0.96	
mixed										
Leu−Asn	0.20 ± 0.11	0.85 ± 0.18	−0.49 ± 0.19	−0.46 ± 0.19	0.88 ± 0.20	0.51 ± 0.41	0.46 ± 0.38	−0.02 ± 0.10	0.25 ± 0.26	−0.32 ± 0.19
Leu−Gln	0.15 ± 0.10	0.41 ± 0.14	−0.29 ± 0.09	−0.38 ± 0.09	0.50 ± 0.14	0.33 ± 0.14	−0.18 ± 0.11	−0.02 ± 0.10	−0.18 ± 0.08	−0.24 ± 0.08
Leu−Lys	−0.33 ± 0.07	−0.12 ± 0.09	0.15 ± 0.11	0.17 ± 0.12	0.14 ± 0.12	0.48 ± 0.16	0.60 ± 0.19	−1.88 ± 0.05	−0.38 ± 0.07	−0.12 ± 0.08
Asn−Lys	0.23 ± 0.12	0.08 ± 0.12	−0.01 ± 0.11	0.07 ± 0.11	−1.21 ± 0.06	−1.49 ± 0.06	−1.58 ± 0.05	−3.80 ± 0.05		
Met−Ser	−0.17 ± 0.09	0.05 ± 0.11	0.16 ± 0.11	0.14 ± 0.11	−0.02 ± 0.10	−0.02 ± 0.11	0.72 ± 0.21	−0.37 ± 0.08	−0.32 ± 0.07	−0.59 ± 0.08
apolar										
Leu−Leu	−1.19 ± 0.04	−0.88 ± 0.03	−0.37 ± 0.09	−0.34 ± 0.09	−0.19 ± 0.08	−0.37 ± 0.08	−0.52 ± 0.08	0.11 ± 0.11	−0.29 ± 0.07	−0.24 ± 0.10
Val−Val	−1.02 ± 0.05	−0.50 ± 0.07	−0.10 ± 0.10	−0.11 ± 0.09	−0.07 ± 0.09	−0.64 ± 0.06	−0.18 ± 0.10	0.11 ± 0.11	−0.06 ± 0.08	−0.11 ± 0.10
Leu−Val	−1.04 ± 0.05	−0.72 ± 0.05	−0.16 ± 0.10	−0.35 ± 0.09	−0.04 ± 0.09	−0.17 ± 0.09	0.00 ± 0.11	0.11 ± 0.11	−0.03 ± 0.08	−0.29 ± 0.10
aromatic										
Trp−Trp	−1.66 ± 0.06	−1.73 ± 0.06	−1.40 ± 0.05	−1.28 ± 0.05	−0.40 ± 0.11	−1.09 ± 0.06	0.56 ± 0.18	−0.81 ± 0.10	−1.21 ± 0.05	−1.49 ± 0.06
Tyr−Tyr	−1.07 ± 0.08	−1.25 ± 0.07	−0.92 ± 0.07	−1.08 ± 0.06	−0.13 ± 0.13	−1.42 ± 0.06	0.41 ± 0.18	−0.95 ± 0.09	−3.93 ± 0.07	−2.27 ± 0.06
Phe−Phe	−2.06 ± 0.05	−1.81 ± 0.06	−0.49 ± 0.08	−0.79 ± 0.07	−0.08 ± 0.14	0.35 ± 0.18	−0.05 ± 0.14	−0.52 ± 0.59	−1.84 ± 0.05	−0.53 ± 0.08
His−Phe	−0.99 ± 0.08	−0.88 ± 0.09	−0.66 ± 0.07	−0.71 ± 0.07	0.07 ± 0.16	−0.19 ± 0.16	−0.36 ± 0.09	−0.83 ± 0.09	−6.16 ± 0.18	−1.25 ± 0.07
Trp−Tyr	−1.56 ± 0.06	−1.32 ± 0.07	−1.05 ± 0.06	−1.11 ± 0.06	−0.33 ± 0.11	−0.63 ± 0.10	−0.43 ± 0.10	−1.02 ± 0.09	−2.38 ± 0.04	−1.47 ± 0.07
backbone										
BB−BB	0.23 ± 0.13	0.35 ± 0.15	−0.37 ± 0.08	0.33 ± 0.13	−0.81 ± 0.07	−1.38 ± 0.07	−0.23 ± 0.20	−1.50 ± 0.10	−4.04 ± 0.08	−1.66 ± 0.07

[a]Pairs have been grouped according to chemical properties: charged, apolar, polar, aromatic, mixed, and backbone. Statistical errors were obtained using Bayesian statistics. Only dimerization free energies obtained from free simulations are shown.

dx.doi.org/10.1021/ct200599d | J. Chem. Theory Comput. 2012, 8, 1003−1014

In all simulations, the temperature was kept close to its target value, 298 K, using Berendsen's weak coupling algorithm (τ_T = 0.1, 0.1, and 0.3 ps for AA, UA, and CG, respectively). The SCAs and the solvents were coupled separately. The pressure was maintained close to 1 bar using an isotropic weak coupling scheme (τ_P = 1.0, 1.0, and 3.0 ps for AA, UA, and CG, respectively).[28] The integration time step was 2 fs for both AA and UA simulations and 30 fs for the CG systems. All bonds in the AA and UA systems, as well as the bonds of the aromatic residues in the CG systems, were constraint with the LINCS algorithm.[29]

2.3. Calculation of Dimerization Free Energies. The free energy of dimerization, ΔG^{dimer}, was calculated using the formula derived for two particle systems:[30]

$$\Delta G^{\text{dimer}} = \Delta G^{\varnothing} = -k_B T \ln\left(\frac{v_d n_1}{v^{\varnothing} n_0}\left[\frac{v}{v_d} - 1\right]\right) \quad (1)$$

where k_B is Boltzmann's constant, T is the temperature, v_d is the dimer volume (calculated as the sum of the experimental SC volumes[31]), $v^{\varnothing}$ is the standard volume of 1.66 nm^3 (equivalent to a concentration of 1 mol/L), v is the volume of the system, and n_0 and n_1 are the respective numbers of monomers and dimers counted in the trajectory. The discrimination between the bound and unbound states was based on the distance between the center of mass (COM) of the SCs. A SC pair was considered in a bound state if the distance between their COMs was less than a cutoff. The cutoff was different for each pair, chosen to equal the distance of the first minimum of the SC−SC radial distribution function (RDF). Most simulations showed hundreds to thousands of association and dissociation events. Errors are calculated using the Bayesian statistics approach as previously described.[30]

In order to calculate statistically reliable free energy of dimerization, at least 50 association and dissociation events are needed. For the amino acid SC pairs that did not show sufficient spontaneous events, a potential of mean force (PMF) was calculated as a function of the SC's COM distance. Simulations were run for distances constrained in the range 0.24−1.50 nm with a 0.02 nm interval. At each distance, the system was simulated for 2 ns, from which the first 500 ps were discarded as an equilibration period. Over the remaining simulation time, the mean constraining force was calculated using the constraint pulling code implemented in GROMACS and integrated as described by Hess et al.[32]

2.4. Analysis of Structural Binding Modes. The structural modes of dimerization of the SCAs were analyzed on the basis of their relative orientation at short distances (<0.75 nm). The SCAs' relative orientation was described by a set of simple geometrical reaction coordinates described here. The distance between the centers of mass of the SCAs was used in all cases. For aromatic side chains, both the angle between the normals of the ring planes (*"angle"*) and the dihedral angle between two vectors through SCA rings (*"dihedral"*) were used. The ring plane was defined by the triangle of three atoms in the ring. The dihedrals were defined using two atoms on each ring. For nonaromatic SCAs, we used the angle between vectors describing the long axis of the molecules, where the vector was defined by two atoms of the side chain. Both the angle and the dihedral were plotted against the distance between the SCAs. The probability distributions for each reaction coordinate were calculated, normalized, and converted into free energy surfaces using

$$\Delta G = -k_B T \ln(P) \quad (2)$$

where P is the normalized probability. Analyses were performed using standard GROMACS analysis tools and local scripts. Visual analysis of the trajectories was performed using VMD.[33]

2.5. Calculation of Complex Lifetimes. The lifetimes of the bound state of amino acid SC pairs was calculated using the following time-correlation function also called the survival correlation function:[34]

$$S_L(t) = \frac{1}{(t_{\text{tot}} - \Delta t)} \sum_{t=0}^{t_{\text{tot}}-\Delta t} p(t, t + \Delta t) \quad (3)$$

where $p(t,t + \Delta t)$ is set to 1 if the SCs are bound for a period of time Δt after their initial dimerization at time t and t_{tot} is the length of the trajectory analyzed. To speed up the analysis, the 240 ns trajectories were split into 50 blocks of 4.8 ns, which is still about 2 orders of magnitude longer than the observed lifetimes (see Results). A running average of 5 ps was applied to filter out fast association−dissociation events. The lifetime of a complex, τ_L, was obtained by a weighted (with the standard deviation) least-squares fit of the mean of $S_L(t)$ over the 50 blocks using the following expression:

$$\langle S_L(t) \rangle \approx \exp[-(t/\tau_L)] \quad (4)$$

3. RESULTS AND DISCUSSION

We present the analysis of the dimerization behavior of a representative set of amino acid side chain pairs. Table 2 lists the complete set of dimerization free energies, ΔG^{dimer}, calculated for 21 pairs, in three different solvents (water, *n*-octanol, and decane) and for three different force fields (OPLS, GROMOS, and MARTINI). For pairs that showed insufficient dimerization events, PMFs were determined, and the depth of the well at the contact distance of these PMFs is reported instead (Table 3). The results are presented and discussed in more detail in the following sections. First, the AA OPLS FF is

Table 3. Deepest Well Points Obtained from Potentials of Mean Force, Given in $k_B T$[a]

	octanol			decane		
pair	MARTINI	GROMOS	OPLS	MARTINI	GROMOS	OPLS
Asn/Asn				−2.7 ± 0.1	−7.5 ± 0.5	−11.7 ± 0.1
Gln/Gln				−2.7 ± 0.1	−8.0 ± 0.2	−9.2 ± 0.2
Gln/Asn				−2.8 ± 0.1	−7.8 ± 0.2	−10.1 ± 0.6
Asn/Lys				−4.5 ± 0.1	−24.6 ± 0.2	−29.3 ± 0.3
Arg/Asp	−5.0 ± 0.1	−19.2 ± 1.1	−18.8 ± 2.2	−11.7 ± 0.1	−118.7 ± 0.2	−134.3 ± 0.2
Lys/Glu	−4.5 ± 0.1	−18.5 ± 1.7	−17.0 ± 1.9	−11.2 ± 0.1	−127.6 ± 0.2	−149.8 ± 0.3

[a]PMFs were only calculated for amino acid pairs that showed insufficient association−dissociation events in the free simulations.

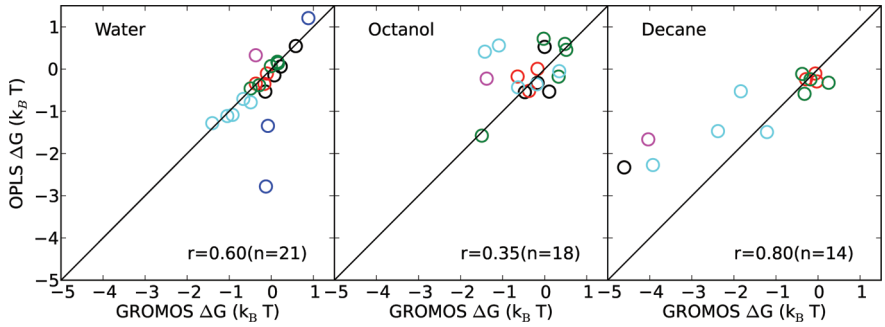

Figure 1. Comparison of dimerization free energies for the OPLS and GROMOS FFs in water (left panel), *n*-octanol (middle panel), and decane (right panel). Blue circles represent charged SC pairs; black circles, polar SC pairs; red circles, apolar SC pairs; green circles, mixed SC pairs; cyan circles, aromatic SC pairs; and magenta circles, the backbone−backbone pair. The black line represents a regression of 1. *r* is the Pearson correlation coefficient, calculated over *n* data points.

compared to the UA GROMOS FF, followed by a comparison of both atomistic FFs to the CG MARTINI FF. The modes of binding are structurally characterized. The ΔG^{dimer} values obtained with the three FFs are then compared to contact energies obtained from a knowledge-based potential. Finally, the lifetimes of the bound SCA pairs in water for the different FFs are discussed.

3.1. OPLS versus GROMOS. The correlation between the SCAs ΔG^{dimer} obtained using the GROMOS 53A5/53A6 and OPLS FF in the three different solvents is shown in Figure 1. The raw data together with error estimates are listed in Table 2.

In water, the ΔG^{dimer} averaged over the 21 pairs is -0.22 ± 0.51 and -0.42 ± 0.81 k_BT for the GROMOS and OPLS FFs, respectively. The root-mean-square difference (RMSD) between the two FFs is 0.68 k_BT. The oppositely charged amino acids (Arg−Asp and Lys−Glu, blue circles) show much stronger interaction (more negative ΔG^{dimer}) for the OPLS FF. Omitting these two pairs, the RMSD decreases to 0.23 k_BT. The RMSDs for the polar, apolar, aromatic, and mixed SCA groups are 0.24, 0.11, 0.17, and 0.06 k_BT, respectively. The correlation coefficient, *r*, is 0.60 over 21 pairs and 0.92 when omitting the charged pairs (Figure 1).

In *n*-octanol, for the 18 pairs that showed sufficient association−dissociation events, there is a considerable, inter-FF spread of the ΔG^{dimer}. The averaged ΔG^{dimer} values are -0.35 ± 0.63 and -0.09 ± 0.56 k_BT for the GROMOS and OPLS FF, respectively. RMSDs are 0.72 k_BT for all 18 pairs and 0.42, 0.30, 1.12, and 0.41 k_BT for the polar, apolar, aromatic, and charged SC groups. The correlation coefficient over the 18 pairs was $r = 0.35$ (Figure 1). For pairs that did not show sufficient association−dissociation events, the PMF was calculated to characterize the interaction strength. This was the case for Lys−Lys, Arg−Asp, and Lys−Glu pairs. The well's depth in the PMF for the oppositely charged SCA pairs (Arg−Asp and Lys−Glu) is similar for the OPLS and GROMOS FFs (cf. Table 3). For the equally charged pair (Lys−Lys) no well depth could be determined since the PMF does not show a minimum. Both PMFs show a good agreement (data not shown).

Note that simulations of *n*-octanol have shown that it may adopt local structure.[35] These structures are more pronounced in hydrated octanol as compared to dry octanol, which we choose for our simulations. Careful examination of the trajectories showed that no macroscopic structure (micelles or inverse micelles) formed. The local structure was characterized by calculating the oxygen−oxygen and oxygen−hydrogen RDFs, which described a similar structure as shown in Figure 4 of MacCallum et al.[35] (data not shown). As far as we could determine, the local structure observed did not impair the sampling of the SCA−SCA interactions.

In decane, only 14 pairs showed sufficient association−dissociation events, and their averaged ΔG^{dimer} values are -1.80 ± 2.01 and -0.92 ± 0.77 k_BT for GROMOS and OPLS, respectively. The RMSDs are 1.71 k_BT for the 14 pairs and 2.28, 0.16, 2.43, and 0.34 k_BT for the polar, apolar, aromatic, and charged SC groups. The correlation coefficient over the 14 pairs is $r = 0.80$ (Figure 1). The depths of the well in the PMF for the pairs that showed insufficient association−dissociation events in decane (i.e., Lys−Lys, Arg−Asp, Asn−Lys, Gln−Gln, Asn−Asn, Gln−Asn, Arg−Asp, and Lys−Glu) are indicated in Table 3. Two PMFs (Lys−Glu and Lys−Lys) are shown in Figure 3. In the case of the polar pairs (Asn−Asn, Gln−Gln, and Asn−Gln), the wells are considerably deeper for the OPLS FF (∼1−4 k_BT), consistent with the lack of association−dissociation events in the free simulations of these pairs using the OPLS FF. In the case of the oppositely charged and mixed pairs (Arg−Asp, Lys−Glu, and Asn−Lys), a similar trend is observed, although the differences are larger (∼5−22 k_BT). For the equally charged Lys−Lys pair, the interaction is virtually the same in both the GROMOS and the OPLS FF (Figure 3).

The RMSD between the FFs over all pairs that showed sufficient association−dissociation events are 0.68, 0.72, and 1.71 k_BT for water, *n*-octanol, and decane, respectively, and 1.10 k_BT averaged over all three solvents. The fact that their performance is most alike in water (smallest RMSD) is to be expected, given that amino acid parameters have been primarily derived for aqueous solutions. Larger differences between the two FFs are found in *n*-octanol and decane. Especially, the difference in dimerization free energies for aromatic and charged side chains is worrisome, with a RMSD exceeding 2 k_BT. The discrepancy may partly result from the difference in solvent models (absence of aliphatic hydrogens and corresponding partial charges in the GROMOS FF). The interactions between charged SCAs are stronger in the OPLS FF in all solvents. We attribute this to the difference in the distribution of partial charges on charged groups between both FFs. The OPLS FF has larger partial charges when compared to the GROMOS FF, leading to larger dipoles in the molecules and therefore stronger interactions, especially in solvents of low polarity. The aromatic SCs correlate well in water, yet some stronger interaction is observed with the GROMOS FF in both *n*-octanol (Trp−Trp, Tyr−Tyr) and decane (all except Trp−Trp). In these cases, the distribution of charge difference between the

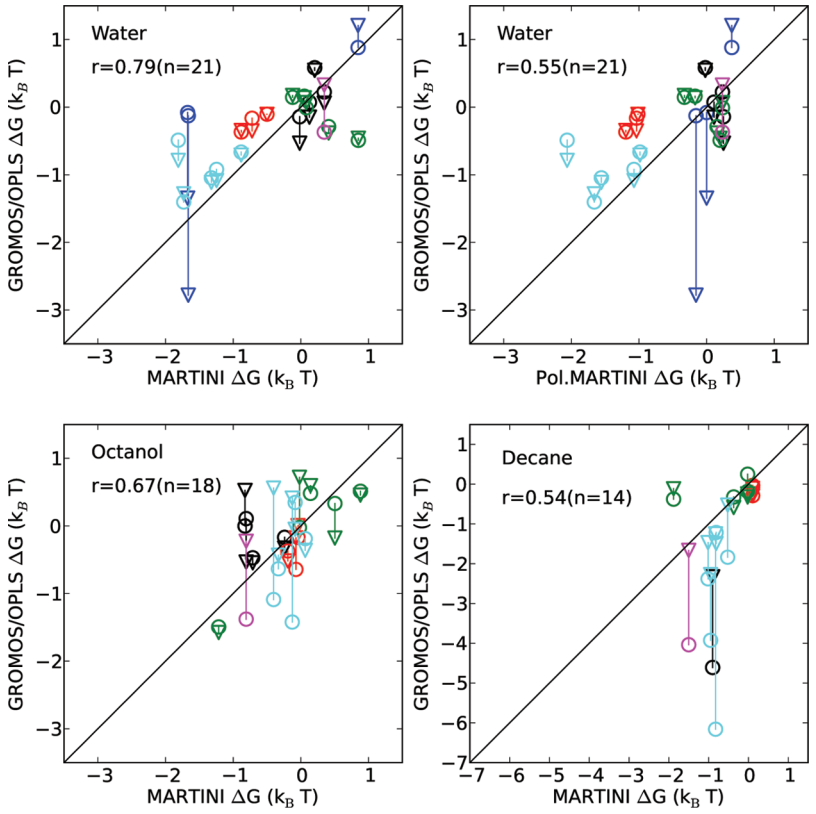

Figure 2. Comparison of dimerization free energies between MARTINI and OPLS or MARTINI and GROMOS FFs in water (top left panel), polarizable water (top right panel), octanol (bottom left panel), and decane (bottom right panel). Circles (○) represent a GROMOS−MARTINI comparison; triangles (▽) represent an OPLS−MARTINI comparison. Lines connect the same pair in each of the two atomistic models. Blue are charged, black are polar, red are apolar, green are mixed, cyan are aromatic SC pairs, and magenta is the backbone−backbone pair. The black line represents a regression of 1. r is the Pearson correlation coefficient, calculated over n SC pairs.

ring carbons and protons is larger in the GROMOS FF than in the OPLS FF, leading to larger local dipoles.

Backbone−backbone interactions are also stronger for the GROMOS FF. This is true in all solvents and, again, most pronounced in decane. Experimental data on the dimerization free energy of the backbone analogue molecule *n*-methylacetamide (NMA) point to an even larger difference between dimerization in water versus organic solvent. Values in the range of 7−9 k_BT are listed by Ben-Tal et al.,[21] compared to a ΔG^{dimer} (water−decane) of 3.7 and 2.0 k_BT for GROMOS and OPLS, respectively. However, the authors also note that the interpretation of the experimental data is far from trivial.

3.2. MARTINI versus Atomistic. Figure 2 shows the correlation plots between the ΔG^{dimer} obtained for amino acid SCAs using the MARTINI CG FF and the joint data obtained for the atomistic FFs discussed above. The raw data and associated error estimates are given in Table 2. To ease the comparison, the results for the atomistic FFs were averaged and are referred to as $\overline{AT}$ in the following.

In water, the average ΔG^{dimer} over the 21 pairs is −0.44 ± 0.85, −0.42 ± 0.81, and −0.22 ± 0.51 for the MARTINI, OPLS, and GROMOS FFs, respectively. The RMSD over these pairs and between the MARTINI FF and $\overline{AT}$ is 0.54 k_BT. It should be noted that for the charged residues there is a large spread between OPLS and GROMOS (cf. the connecting lines in Figure 2). Omitting the charged SC yields a RMSD value of 0.52 k_BT. The small difference reflects a cancellation between the GROMOS and OPLS FFs, with the average dimerization free energy close to the value obtained for the MARTINI FF. The groups of polar, apolar, aromatic, and mixed SCAs have RMSDs of 0.27, 0.46, 0.58, and 0.69 k_BT, respectively. The Pearson correlation coefficient over all pairs is r = 0.79 (Figure 2). The average ΔG^{dimer} over the 21 pairs is not affected when switching to the polarizable MARTINI water model: −0.45 ± 0.74. The RMSDs change to 0.66 k_BT for all pairs and 0.42, 0.84, 0.70, and 0.46 k_BT for polar, apolar, aromatic, and mixed SCAs, respectively. The correlation of the dimerization free energies with respect to the AA models becomes slightly worse (Pearson coefficient drops to 0.55, Figure 2); however, this is largely attributed to the effect of the charged pairs. Whereas with standard MARTINI water the $\overline{AT}$ value is closely matched, the polarizable model matches GROMOS very closely, but not OPLS.

In *n*-octanol, the averaged ΔG^{dimer} values over the 18 pairs showing sufficient association−dissociation events are −0.22 ± 0.49, −0.09 ± 0.56, and −0.35 ± 0.63 k_BT for MARTINI, OPLS, and GROMOS FFs, respectively. The RMSDs between MARTINI and $\overline{AT}$ are 0.40 k_BT for all pairs and 0.63, 0.24, 0.27, and 0.38 k_BT for the polar, apolar, aromatic, and mixed SCAs, respectively. The correlation coefficient over these 18 pairs is r = 0.67 (Figure 2). For the charged pairs, insufficient sampling was achieved to determine ΔG^{dimer}. The PMFs, however, show that the MARTINI FF systematically underestimates the attraction between unlike charges when compared to both atomistic FFs, as evidenced by the well depths of Lys−Glu and Arg−Asp pairs (Table 3). Similarly, the repulsion between like charges is underestimated (Lys−Lys pair, data not shown).

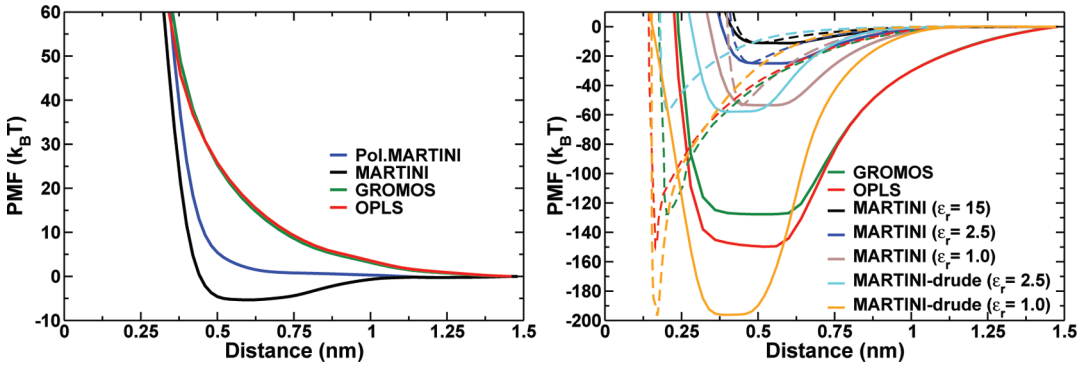

Figure 3. Potential of mean force for Lys−Lys (left) and Lys−Glu (right) SC pairs in decane. Solid lines are PMF plotted against the center of mass distance; dashed lines are PMFs plotted against charge group minimum distance. Black lines are MARTINI, green lines are GROMOS, red lines are OPLS, blue lines are polarizable water MARTINI (ε_r = 2.5), brown lines are MARTINI without dielectric screening (ε_r = 1.0), cyan lines are an alternative model for charged residues in MARTINI (drude-like, see text) with ε_r = 2.5, orange lines are the same model with ε_r = 1.0. Errors do not exceed the thickness of the lines.

In decane, for the 14 pairs that showed sufficient association−dissociation events, the averaged ΔG^{dimer} values are −0.61 ± 0.61, −0.92 ± 0.77, and −1.80 ± 2.01 k_BT for the MARTINI, OPLS, and GROMOS FFs, respectively. The RMSDs are 1.36 k_BT for all pairs and 2.57, 0.29, 1.70, and 0.82 k_BT for the polar, apolar, aromatic, and mixed SCAs, respectively. Over these 14 pairs, the correlation coefficient is r = 0.54 (Figure 2). The PMFs calculated for the charged SCA pairs indicate much weaker Coulombic interactions in the MARTINI FF than in the atomistic ones (Table 3). The examples of the Lys−Glu and Lys−Lys pairs are shown in Figure 3. The Lys−Lys pair exhibits a slightly attractive interaction with the MARTINI FF, while the atomistic FFs are purely repulsive. The behavior of the charged SCA interactions in the MARTINI FF is improved with the polarizable version of the water model, also shown in Figure 3. The depth of the well of the Lys−Glu pair PMF decreases from −11 to −25 k_BT, albeit still much smaller than the −120/−140 k_BT value when using the atomistic models. In the case of the Lys−Lys pair, the attractive range observed with the standard MARTINI water disappears with the polarizable model.

The RMSDs of ΔG^{dimer} between the MARTINI FF and $\overline{AT}$ are 0.54, 0.40, and 1.36 k_BT for water, *n*-octanol, and decane, respectively, and 0.86 k_BT averaged over the three solvents. Comparing the MARTINI FF used with polarizable water to $\overline{AT}$ in water gives a RMSD of 0.65 k_BT. Thus, the difference between MARTINI and GROMOS/OPLS is on the order of k_BT, comparable to the difference between the two atomistic FFs. Main outliers to this last observation are the interaction between Phe side chains in water, which is more attractive in the MARTINI FF, and interactions involving charged and polar SCAs in the apolar media.

In the case of polar SCs, this discrepancy is due to the lack of an explicit representation of the polarity. The introduction of partial charges, in line with the polarizable water model,[27] may improve this behavior.

For the charged residues, the large difference observed between the atomistic and CG approaches results from two factors. First, in MARTINI, a high implicit dielectric screening (ε_r = 15) is used, while in the atomistic models no dielectric screening is used (ε_r = 1). Reducing the dielectric screening used in MARTINI simulations does increase the interaction strength between the charged SCs (See Figure 3): (i) in the polarizable MARTINI model (ε_r = 2.5), the interaction strength is increased by a factor of 2; (ii) when no dielectric screening is used (ε_r = 1.0), the interaction strength increases further by another factor of 2 (not shown). Second, the size of the CG beads (defined by the van der Waals potential) limits the approach of the charged beads to ∼0.5 nm, whereas the charges in the atomistic models can come much closer. Because the Coulombic interactions fall off as $1/r$, the total interaction strength between the charged atomistic side chains is also affected and is much more attractive. This effect becomes evident when the PMFs are plotted against the minimum distance between the charges instead of the COM distance of the SCAs (dashed lines in Figure 3). To increase the binding strength in the MARTINI force field, we designed an alternative model for charged side chains in which the electrostatic and van der Waals interactions are carried by two different particles connected by a constrained bond of length 0.11 nm (drude-like). The van der Waals center and the charge both have a mass of 36 amu. Using this setup, the charges may approach each other much more closely, and the interaction increases by approximately a factor of 5 for the lysine/glutamate pair as compared to normal MARTINI (cf. Figure 3). Combining this approach with an absence of implicit dielectric screening (ε_r = 1) increases the interaction strength by a factor 18 compared to the normal MARTINI topology. The latter setup yields SCA pair interactions stronger than when using the GROMOS and OPLS force fields. This is due to the spread of the charge over multiple atoms in the atomistic force field.

3.3. Structural Analysis. To further characterize the binding of SCAs in different force fields and solvents, we analyzed the structural binding modes of the SCA pairs. To do so, we build 2D probability distributions correlating different reaction coordinates. For the aromatic pairs, both the angle between the ring normals (*"angle"*) and the dihedral angle between two axes within the ring planes (*"dihedral"*) are correlated with the COM distance. For the nonaromatic pairs, the angle between the SCAs' long axes was correlated with the COM distance. In the MARTINI FF, only the aromatic pairs have sufficient beads to define the angle and dihedral. Therefore, for the other pairs, only OPLS and GROMOS could be compared. The analysis of the angles was combined with a careful visual inspection of the trajectories. The results are shown in Figures 4 and 5 for the case of the Phe−Phe pair; plots for the other pairs can be found in the Supporting Information, Figures 1−22.

dx.doi.org/10.1021/ct200599d | J. Chem. Theory Comput. 2012, 8, 1003−1014

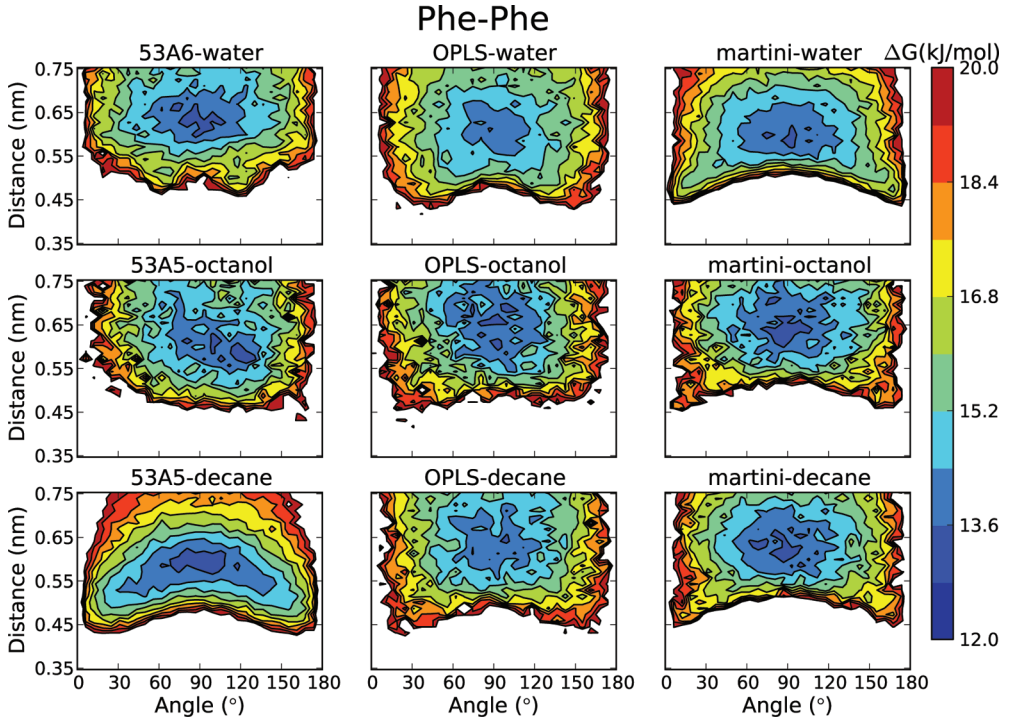

Figure 4. Two-dimensional histograms showing the free energy profiles of the interactions between two phenylalanine side chains in different solvents and different force fields. The angle between the normal of the ring plains is plotted against the center of mass distance. The normal of the ring plains is defined by C_γ, $C_{\varepsilon 1}$, and $C_{\varepsilon 2}$ for GROMOS and OPLS and all three ring beads for MARTINI. Colors indicate free energy levels in kJ/mol.

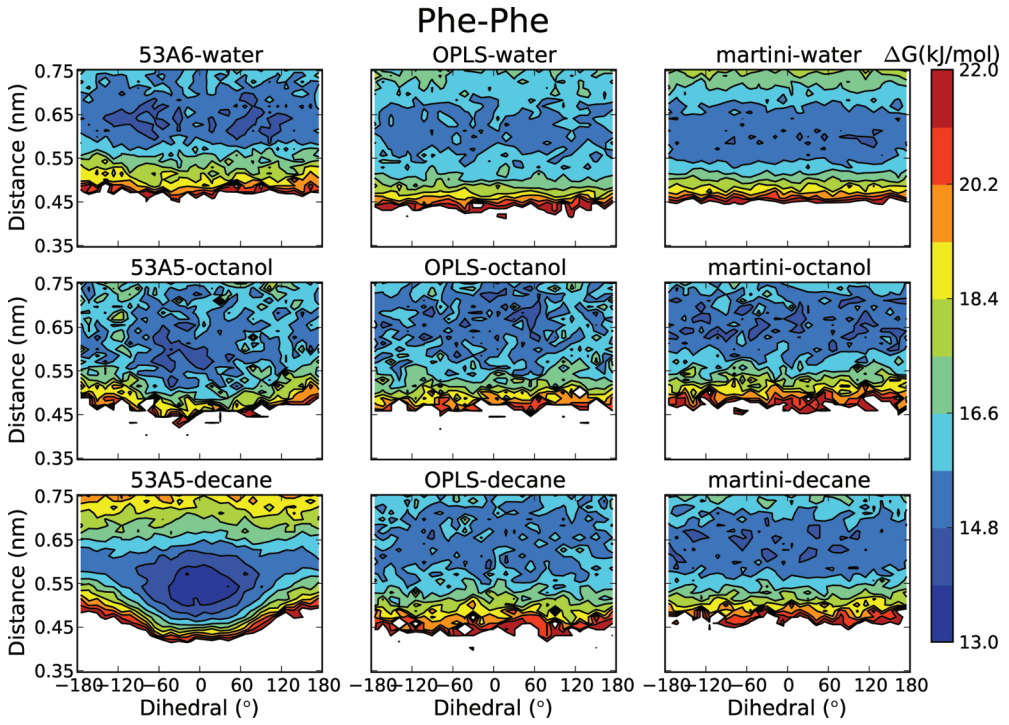

Figure 5. Two-dimensional histograms showing the free energy profiles of the interactions between two phenylalanine side chains in different solvents and different force fields. The dihedral angle between the vectors in the plane of the ring is plotted against the distance between the centers of mass of the SCAs. The vectors are defined by C_γ and C_ζ for GROMOS and OPLS and the first two ring beads for MARTINI. Colors indicate free energy levels in kJ/mol.

Phenyl rings in general and phenylalanine residues in particular are known to bind in mainly three different modes: T-stacking, straight-stacking, and shifted-stacking. T-stacking and shifted-stacking are most abundant in proteins.[36] For water and octanol, Figures 4 and 5 show some typical characteristics that are seen in all of the aromatic pairs. First, the lower free energy around a 90° angle is due to entropic reasons: more conformations are possible for a perpendicular orientation as compared to

parallel orientations. Second, the position of the free energy minimum in the distance dimension corresponds to the first minimum of the radial distribution function. And third, the lower free energy at low distances and angles around 0 and 180° is due to the flat stacking of the rings that is preferred at small distances. Except for those general characteristics, in water and octanol no specific binding mode is found in either of the force fields. In decane, GROMOS shows a strong preference for the shifted stacking, where the methyl groups are located straight above the ring of the other side chain. In Figure 5, this orientation corresponds to the lower free energy at 0°. A cluster analysis showed that this conformation accounted for ∼40% of the bound state. For OPLS in decane no strongly specific orientation was observed. For MARTINI, orientations in all three solvents are governed by optimal packing of the LJ spheres, leading to a preferred binding in the ridges between the coarse-grain beads.

For GROMOS and OPLS, two tyrosine rings in water are often in a stacked conformation, where there is no preferred orientation for the ring substituents (Supporting Information (SI), Figures 1 and 2). In decane, the interaction is mainly guided by the hydrogen bond between the alcohol groups for OPLS and GROMOS. GROMOS has a second conformation at a closer distance in which the alcohol group points to the middle of the ring of the second molecule. In octanol, this latter conformation is observed for both GROMOS and OPLS, however, less pronounced. Using MARTINI, in decane and octanol, the rings are mainly found in a stacked conformation. In water, orientations are again governed by sphere packing.

The histidine−phenylalanine pair has no specific binding mode in water for any of the force fields (SI, Figures 3 and 4). In decane, OPLS and GROMOS favor a conformation where the ε_2 hydrogen is pointing to the middle of phenyl ring, leading to a 90° angle between the rings. This orientation was very strongly preferred for GROMOS compared to OPLS. The difference is less pronounced in octanol. For MARTINI, orientations in all three solvents are governed by the packing of the CG beads.

The tryptophan−tryptophan pair preferably binds with the amine hydrogen of one molecule pointing to the six ring of the second molecule in both GROMOS and OPLS FFs (SI, Figures 5 and 6). This orientation is strongly present in decane and, for OPLS, in octanol. In water, no specific binding mode is observed. In MARTINI, there is a specific binding mode where the rings stack parallel, in that way maximizing the total contact surface.

Using GROMOS and OPLS, the Trp−Tyr pair has two main conformations: either the Tyr alcohol group is pointing toward the middle of the Trp five ring or the Trp amine group is pointing toward the middle of Tyr six ring (SI, Figures 7 and 8). In OPLS, the former conformation is more abundant (equivalent to the 90° dihedral angle, see Supporting Information), while in GROMOS the latter is (equivalent to the ±180° dihedral angle). In water, there is no preferred orientation, except for small distances (<0.5 nm) where the rings are mainly found in a stacked orientation. This orientation is preferred for MARTINI in all three solvents.

For the four polar pairs (Ser−Ser, Asn−Asn, Gln−Gln and Gln−Asn) OPLS and GROMOS behave very similarly in all solvents (SI, Figures 9, 10, 11, and 12). Interactions in octanol and decane are governed by hydrogen bonds. In decane, this effect is the most pronounced, leading to only a small number of structural arrangements being sampled and a strong binding energy (Tables 2 and 3). For the Ser−Ser pair, the hydrogen bond is between the alcohol groups, which leads to an average angle between the molecules around 90°. For the other polar pairs, a molecule angle around 180°, corresponding to the hydrogen bond between the oxygen and either of the amine hydrogens, is most prevalent. In octanol, the partial charges are more strongly screened, and the interactions are consequently weaker and less specific, although the hydrogen bond is still most prevalent. The pairs involving Asn or Gln also sample the hydrogen bond between both amine groups. In water, the screening of charges is even stronger, and no specific binding is observed.

All three pairs of apolar molecules as well as the mixed pairs (Leu−Asn, Leu−Gln, Leu−Lys, and Met−Ser) show no specific binding mode in any of the solvents when using a high resolution force field (SI, Figures 13−19). The angles wobble around 90° due to entropic reasons. Valine−valine has a slightly smaller minimum distance when in contact.

The charged-polar mixed pair Asn−Lys has a strongly preferred binding mode where one of the (charged) amine protons forms a hydrogen bond with the Asn oxygen (SI, Figure 20). The two minima found in the plot result from the presence of two conformations of the lysine carbon tail (either stretched or U-shaped).

The charged Lys−Glu pair in decane and octanol always binds via its charged groups. A large range of angles and the associated change of COM−COM distance is sampled and results from the change of conformation of the carbon tails of both molecules (SI, Figure 21). In water, using the OPLS force field, the side chains show the same mode of binding, although less pronounced. In GROMOS, potentially due to the downscaling of the charges in 53A6 compared to the 53A5 version there is no specific binding mode in water.

For the charged Arg−Asp pair, there are strong differences between force fields and solvents (SI, Figure 22). In decane, using OPLS, the charged Asp oxygens are strongly coordinated by the η_1 and η_2 hydrogens of Arg, while in octanol they are coordinated by η_2 and ε hydrogens. In water, both conformations are found. In decane, using 53A5, Asp binds between the η_2 and ε carbons of Arg, with the planes of the molecules being perpendicular. In octanol, Asp binds Arg at both η_1/η_2 and η_2/ε positions, with the molecule planes primarily in a perpendicular orientation. In water, using 53A6, no specific binding mode is observed.

3.4. Comparison to Knowledge-Based Potentials. Unfortunately, there is no direct experimental data that could be used to compare the calculated ΔG^{dimer} against. There is, however, a long history of extracting side chain−side chain contact energies for amino acid pairs from their occurrence in protein structures.[37−39] These contact energies have been used to derive energy functions for CG FFs, the so-called knowledge-based potentials (KBP),[40] as well as to test existing FFs.[41] Although the environment of an amino acid buried in a protein is different than free in solution (e.g., different orientational freedom and screening of electrostatics) and the methods used to extract such KBPs are not flawless,[20] these contact energies provide us with a useful independent benchmark. We use the set of contact energies obtained by Miyazawa and Jernigan[22] (MJ). It was extracted from a set of 20 000 SC contacts from which a delicate balance of terms was carefully derived and corrected for potential biases.

The comparison of the dimerization free energies obtained from our simulations to the contact energies obtained by MJ is shown in Figure 6. The correlation coefficients, r, between the data from MJ and the OPLS, GROMOS, regular and polarizable MARTINI FFs, and over 20 SCA pairs are 0.48, 0.67, 0.72, and 0.87, respectively. Note that in MJ, the backbone was not

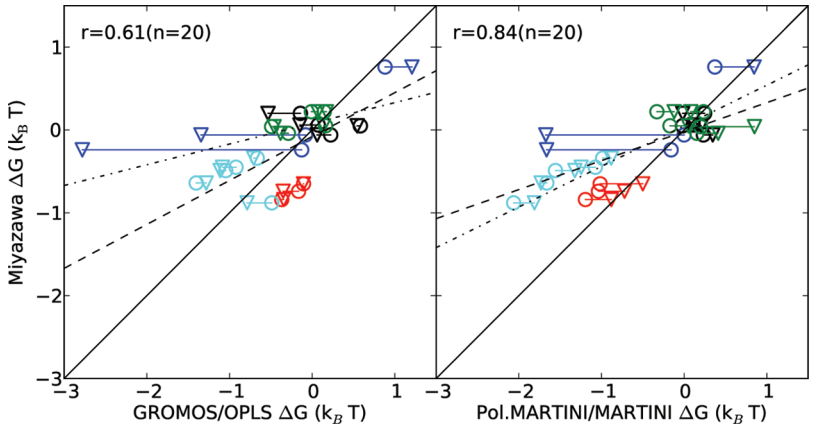

Figure 6. Comparison of calculated SC dimerization free energies to knowledge-based potentials (KBP) derived by Miyazawa and Jernigan.[22] In the left panel, KPB is compared to the atomistic FFs GROMOS (circles (○)) and OPLS (triangles (▽)). In the right panel, KPB is compared to the CG FFs MARTINI (circles (○)) and polarizable MARTINI (triangles (▽)). Lines connect equal SC pairs. Blue are charged, black are polar, red are apolar, green are mixed, and cyan are aromatic SC pairs. The black line represents a regression of 1. Dashed and dashed-dotted lines represent regression fits for GROMOS and OPLS (left panel) and MARTINI and polarizable MARTINI (right panel). r is the Pearson correlation coefficient, calculated over n data points of the average of GROMOS and OPLS (left panel) and of MARTINI and polarizable MARTINI (right panel).

considered. The slope, a, of a linear regression fit of the MJ data versus the GROMOS, OPLS, and regular and polarizable MARTINI FFs data was 0.53, 0.25, 0.35, and 0.49, respectively. The correlation and regression coefficients are in line with those found by Betancourt et al.[20] in their comparison of the GROMOS 43A1 FF to the data from MJ (r = 0.55 and a = 0.30, fitted including all possible amino acid pairs). It is also notable that the correlation with an earlier set of contact energies derived by MJ[40] is much lower (data not shown), suggesting that the additional corrections were valuable.

Overall, the two MARTINI FFs compare better to the MJ SC−SC contact energies than the atomistic ones. Notably, the OPLS and the polarizable MARTINI FFs have stronger unlike charge−charge interactions in comparison to the data from MJ. The correlation coefficients become 0.70 and 0.88, respectively, when these interactions are omitted from the analysis.

In all four FFs, the interactions between aromatic SCs are stronger than those obtained by MJ, in line with the observation that values of interaction energies below -1 k_BT (the case for the rings) estimated by the method used by MJ might be underestimated (too weak).[20] The deviation of these side chains primarily leads to regression coefficients smaller than 1 for all FFs, as was also found by Betancourt et al. It is interesting to note that no significant configurational preference for ring−ring interactions was found in our simulations. This is in great contrast with first principle (ab initio) calculations and protein structure analysis. This suggests that ring−ring interactions in an aqueous solvent might be less specific than being buried in a protein and/or in a nonpolar environment. It was also notable that in most cases simulated the SC−SC interactions experienced an increase of conformational sampling with the increase of the polarity of the solvent (from decane to water).

Considering that the KBP was obtained from structures of soluble proteins,[22] the choice to compare to ΔG^{dimer} obtained from simulations in water seems reasonable. However, the dielectric environment inside a protein is not the same as in aqueous solution. We therefore also calculated the correlation coefficients between the KBP and the ΔG^{dimer} calculated in n-octanol and decane for the different FFs. The correlation coefficients were systematically lower than those obtained in water and always below 0.2 (data not shown).

3.5. Kinetics of Side Chain Association. To assess the variations between the FFs with respect to the association kinetics of the various SCAs, the lifetimes of the SCA bound complex, τ_L, were computed using eqs 2 and 3. Only the aqueous solvent case was considered, for which the results are listed in Table 4. The data show that, for most SCA pairs, the

Table 4. Complex Lifetimes of Amino Acid Side Chain Complexes in ps[a]

pair	pol. MART.	MARTINI	GROMOS	OPLS
charged				
Lys/Glu	24.3 ± 0.7	68 ± 2	12.9 ± 0.3	(1.5 ± 0.2) × 10^2
Lys/Lys	12.8 ± 0.3	9.9 ± 0.8	20 ± 3	18 ± 2
Arg/Asp	38 ± 2	68 ± 1	12.2 ± 0.2	(7 ± 5) × 10^2
polar				
Ser/Ser	20.9 ± 0.6	16 ± 1	12.8 ± 0.5	9.6 ± 0.2
Gln/Asn	19.5 ± 0.3	24 ± 2	16 ± 4	16 ± 3
Asn/Asn	22 ± 1	16.1 ± 0.6	16.0 ± 0.7	20 ± 1
Gln/Gln	18.7 ± 0.6	25 ± 2	23.1 ± 0.9	28 ± 3
mixed				
Leu/Asn	14.3 ± 0.5	8.4 ± 0.4	17 ± 1	16 ± 2
Leu/Gln	14.2 ± 0.2	13.4 ± 0.7	23.6 ± 0.7	31 ± 2
Leu/Lys	18.5 ± 0.4	16.1 ± 0.8	18.3 ± 0.6	21.6 ± 0.6
Asn/Lys	17.0 ± 0.3	20 ± 1	20.9 ± 0.7	17.4 ± 0.4
Met/Ser	17.1 ± 0.2	13.3 ± 0.3	15.3 ± 0.3	15.8 ± 0.7
apolar				
Leu/Leu	25.7 ± 0.1	20.3 ± 0.1	27.5 ± 0.8	27.3 ± 0.8
Val/Val	29.5 ± 0.5	17.6 ± 0.5	21.2 ± 0.4	19.7 ± 0.8
Leu/Val	28.7 ± 0.5	16.5 ± 0.5	24 ± 1	27.8 ± 0.8
aromatic				
Trp/Trp	124 ± 7	110 ± 9	65 ± 2	58 ± 2
Tyr/Tyr	71 ± 4	67 ± 3	56 ± 4	52 ± 2
Phe/Phe	102 ± 2	84 ± 3	36 ± 2	40 ± 2
His/Phe	67 ± 4	56 ± 4	34 ± 1	33.0 ± 0.8
Trp/Tyr	119 ± 8	64 ± 3	53 ± 2	41.0 ± 0.6
backbone				
BB/BB	22 ± 1	16.1 ± 0.6	24.6 ± 0.5	16.9 ± 0.8

[a]Pairs have been grouped according to chemical properties: charged, apolar, polar, aromatic, mixed, and backbone. Statistical errors were obtained as the fit error in the least squares fitting procedure.

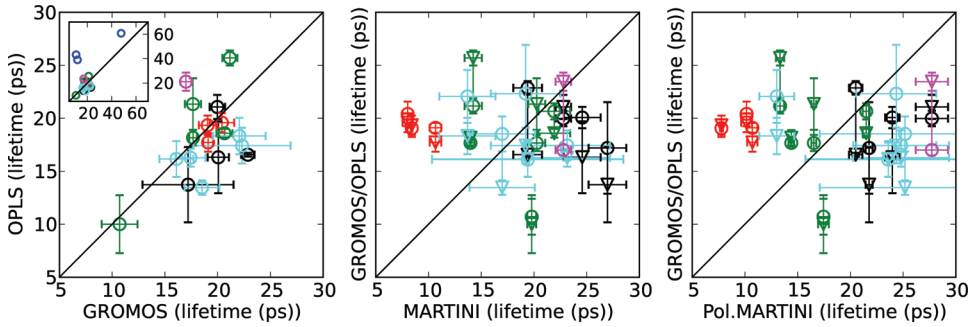

Figure 7. Comparison of residual lifetimes of amino acid SCAs in water between different force fields. In the first pane, circles indicate a SCA pair. In the second and third pane, circles compare GROMOS to MARTINI and triangles compare OPLS to MARTINI. Black represents polar SC pairs, red represents apolar SC pairs, green represents mixed pairs, cyan represents aromatic SC pairs, and magenta represents the backbone−backbone pair. The inset in the first pane shows the charged pairs (blue circles). Charged pairs have been omitted for the second and third plots. Errors in both dimensions are shown as error bars. The black line represents a regression of 1.

lifetime of the complex is on the order of 10−50 ps, irrespective of the nature of the side chains. Since the lifetimes are expected to be proportional to $\exp(-\Delta G^{\text{dimer}})$, the differences between the FFs are largely reflecting differences in dimerization free energy. To appreciate whether there are additional effects, we calculated a residual lifetime defined as $\tau^* = \tau_L/\exp(-\Delta G^{\text{dimer}})$. Figure 7 shows the correlation of these residual lifetimes between the FFs.

The comparison between the residual lifetimes obtained using the OPLS and GROMOS FFs shows a clear correlation (Figure 7). Within the error bounds, for most pairs, the residual lifetimes are the same. The average residual lifetimes for all but the charged pairs are 19 ± 3 ps and 18 ± 4 ps for GROMOS and OPLS, respectively. Thus, for the OPLS and GROMOS FFs, the variations in ΔG^{dimer} are mainly affecting τ_L.

Comparing residual lifetimes obtained using OPLS and GROMOS with those obtained using MARTINI and polarizable MARTINI, a larger residual spread in the values obtained with the CG FFs is apparent (Figure 7). There is a strong influence of the type of amino acid: the apolar pairs have low residual lifetimes (9 ± 1 ps for MARTINI and 10 ± 1 ps for polarizable MARTINI), the polar pairs have the highest residual lifetimes (23 ± 3 ps for MARTINI and 23 ± 3 ps for polarizable MARTINI), and the mixed and aromatic pairs are intermediate (18 ± 3 ps and 18 ± 3 ps for MARTINI and 17 ± 3 ps and 22 ± 5 ps for polarizable MARTINI).

The reason for this residual spread is not entirely clear, but it probably related to differences in the rates of complex formation.

4. CONCLUSIONS AND OUTLOOK

To assess the performance of both atomistic and coarse-grain force fields, we have determined the dimerization free energies, ΔG^{dimer}, of 21 amino acid side chain analogue (SCA) pairs in solvents with different degrees of hydrophobicity (water, *n*-octanol, and decane). Overall, good correlations were found between the all-atom OPLS and united-atom GROMOS FFs, and notably also with the MARTINI FF, which uses a lower resolution. The charged SCAs proved to be the most variable over the FFs, especially in the low dielectric solvents. This was the case not only when comparing the CG MARTINI FF to the atomistic FFs but also when comparing OPLS and GROMOS. The use of the polarizable MARTINI water model improved the agreement of the charge−charge interactions between the MARTINI FF with the atomistic FFs, while not affecting the other interactions. Concerning the difference between the atomistic FFs, the OPLS FF has stronger Coulomb interactions than the GROMOS FF, leading to larger dimerization free energies in apolar solvents.

The high correlation and deviations found in the FFs contribute to defining the areas where one may be confident in the force fields and the points that need be taken with care and improved. For instance, the structural analysis of the binding modes of the respective SCA pairs points to significant differences in the packing of, especially, aromatic residues between the force fields. One would expect that this will also affect the way aromatic residues pack inside a protein. Furthermore, the differences in interaction energies between charged and polar residues in apolar solvents is a serious point of worry. The choice of force field will have consequences for processes such as membrane poration by charged aggregates and peptides (e.g dendrimers and antimicrobial peptides), the cooperativity of charged amino acids penetrating in membranes, the binding of ligands in hydrophobic pockets inside proteins, and membrane protein complexation. In order for a force field to better cover those kinds of phenomena, membrane proteins should be used in addition to soluble proteins for the validation of the force field.

On the basis of the deviations of the MARTINI FF with respect to the atomistic FFs, a few recommendations could be made for future improvements. The phenylalanine residue pair seems to bind too strongly in the MARTINI FF. The Phe SC should be made slightly more attractive toward water than it is currently. A similar conclusion was recently reached by Singh and Tieleman,[42] comparing the binding of Wimley−White pentapeptides to lipid bilayers. The oppositely charged and polar SCs were found to bind too weakly in the apolar solvents. For the charged SCs, this is slightly improved by using the polarizable MARTINI model, which reduces the screening of Coulombic interactions, and improved further by adding an extra, charge carrying bead to the model. In the case of the polar SCs, it could be advantageous to introduce polarizable particles in the line of the polarizable water model. We are currently testing different parametrizations for both the Phe SC and the polar side chains (Asn, Gln, Thr, Ser) to improve their behavior.

The side chain contact energy pairs obtained by Miyazawa and Jernigan from the occurrence of the SC pairs in a large set of protein structures correlated remarkably well with the ΔG^{dimer} in water obtained with the various FFs. This is even more striking considering the completely different approaches used to derive the potentials. It is worth mentioning that the

comparison with contact energies derived earlier by a different method[40] did not correlate as well as with the current[22] set. The larger deviations observed for the aromatic SCs suggest that they may involve more effects difficult to account for, or that the FFs are not parametrized to the same level of accuracy as the other SCs.

The current work demonstrates the capability of modern large scale computational power to systematically assess the quality of side chain interactions in molecular dynamics force fields and identifies some possible improvements to be made with respect to these interactions.

■ ASSOCIATED CONTENT

Supporting Information

2D histograms for all SCA pairs. This information is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: s.j.marrink@rug.nl.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to acknowledge Lars V. Schäfer for useful discussions. X.P. and S.J.M. acknowledge funding from The Netherlands Organization for Scientific Research through an ECHO grant.

■ REFERENCES

(1) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585−590.
(2) Vendruscolo, M.; Dobson, C. M. *Curr. Biol.* **2011**, *21*, R68−R70.
(3) Clementi, C. *Curr. Opin. Struct. Biol.* **2008**, *18*, 10−15.
(4) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. *Science* **2010**, *330*, 341.
(5) Bowman, G. R.; Voelz, V. A.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2010**, *21*, 4−11.
(6) Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K. *Nat. Phys.* **2010**, *6*, 751−758.
(7) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57−83.
(8) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Annu. Rev. Biophys.* **2008**, *37*, 289−316.
(9) Cellmer, T.; Bratko, D.; Prausnitz, J. M.; Blanch, H. W. *Trends Biotechnol.* **2007**, *25*, 254−261.
(10) Periole, X.; Huber, T.; Marrink, S. J.; Sakmar, T. P. *J. Am. Chem. Soc.* **2007**, *129*, 10126−10132.
(11) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(12) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(13) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. *Eur. Biophys. J.* **2011**, *40*, 843−856.
(14) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(15) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(16) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656−1676.
(17) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(18) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(19) Yang, H.; Elcock, A. *J. Am. Chem. Soc.* **2003**.
(20) Betancourt, M. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 72−85.
(21) Ben-Tal, N.; Sitkoff, D.; Topol, I. A.; Yang, A. S.; Burt, S. K.; Honig, B. *J. Phys. Chem. B* **1997**, *101*, 450−457.
(22) Miyazawa, S.; Jernigan, R. L. *Proteins: Struct., Funct., Bioinf.* **1999**, *34*, 49−68.
(23) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(24) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; pp 331−342.
(25) Garrido, N. M.; Queimada, A. J.; Jorge, M.; Macedo, E. A.; Economou, I. G. *J. Chem. Theory Comput.* **2009**, *5*, 2436−2446.
(26) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451−5459.
(27) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. *PLoS Comput. Biol.* **2010**, *6*, 1−17.
(28) Berendsen, H. J. C.; Postma, J. P. M; Gunsteren, W. F. V.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(29) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(30) de Jong, D. H.; Schäfer, L. V.; de Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C.; Grubmüller, H. *J. Comput. Chem.* **2011**, *32*, 1919−1928.
(31) Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjöström, M.; Wold, S. *J. Med. Chem.* **1998**, *41*, 2481−2491.
(32) Hess, B.; Holm, C.; van der Vegt, N. *J. Chem. Phys.* **2006**, *124*, 164509.
(33) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(34) García, A. E.; Stiller, L. *J. Comput. Chem.* **1993**, *14*, 1396−1406.
(35) MacCallum, J. L.; Tieleman, D. P. *J. Am. Chem. Soc.* **2002**, *124*, 15085−15093.
(36) Castellano, R.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210−1250.
(37) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 945−950.
(38) Miyazawa, S.; Jernigan, R. L. *Macromolecules* **1985**, *18*, 534−552.
(39) Sippl, M. *J. Mol. Biol.* **1990**, *213*, 859−883.
(40) Miyazawa, S.; Jernigan, R. L. *J. Mol. Biol.* **1996**, *256*, 623−644.
(41) Betancourt, M. R.; Omovie, S. J. *J. Chem. Phys.* **2009**, *130*, 195103.
(42) Singh, G.; Tieleman, D. P. *J. Chem. Theory Comput.* **2011**, *7*, 2316−2324.